

DOCUMENT RESUME

ED 223 699

TM 820 846

AUTHOR Mason, William M.; Entwisle, Barbara
TITLE A Better Way to Do Contextual Analysis. A Michigan Comparative Fertility Project Working Paper, No. 82-32.
INSTITUTION Michigan Univ., Ann Arbor. Center for Population Studies.
SPONS AGENCY Andrew W. Mellon Foundation, New York, N.Y.; National Inst. of Child Health and Human Development (NIH), Bethesda, Md.
PUB DATE Sep 82
GRANT 1R01-HD-15730-01
NOTE 35p.; Paper presented at the Annual Meeting of the American Sociological Association (San Francisco; CA, September, 1982).
AVAILABLE FROM Publications/Population Studies Center, 1255 South University Avenue, Ann Arbor, MI, 48104 (\$2.00).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)
EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Bayesian Statistics; Birth Rate; Demography; *Estimation (Mathematics); Least Squares Statistics; *Maximum Likelihood Statistics; *Models; Predictive Measurement; Probability; Statistical Significance
IDENTIFIERS *Contextual Analysis; Restricted Maximum Likelihood Bayes

ABSTRACT

The real problems of contextual analysis concern the conceptualization of contextual effects, the kinds of data with which to estimate them, and the selection and implementation of appropriate statistical techniques. This paper focuses on detection; specifically, an approach to contextual analysis based on the estimation and interpretation of a covariance component model with hyperparameters. The restricted maximum likelihood/Bayes (REML/Bayes) estimation procedure described is based on Mason and Wong (forthcoming). This paper illustrates the specification, estimation, and interpretation of a covariance component model with hyperparameters, using an example concerning systematic variability across countries in the determination of individual fertility within countries. The purpose in developing this example is to make the approach accessible to others interested in contextual analysis, but not in the algebraic and statistical complexities that comprise formal presentations. Multilevel analysis using the covariance component framework is an advance in contextual analysis. Whether contextual effects exist for a given substantive problem is one matter; their estimation is another. Researchers may be hard pressed to find such effects unless they use efficient and appropriate estimation techniques in conjunction with appropriate theoretical reasoning. This illustration indicates that the REML/Bayes estimation procedure extracted results which otherwise would have been invisible. (Author/PN)

A Michigan Comparative Fertility Project Working Paper

A BETTER WAY TO DO CONTEXTUAL ANALYSIS*

By

William M. Mason

and

Barbara Entwisle

September, 1982

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

X This document has been reproduced as
received from the person or organization
originating it.

Minor changes have been made to improve
reproduction quality.

• Points of view or opinions stated in this docu-
ment do not necessarily represent official NIE
position or policy.

William M. Mason is Professor of Sociology at the University of
Michigan and Research Associate at the Population Studies Center.

Barbara Entwisle is Assistant Professor of Sociology at Dartmouth
College.

*This paper was presented at the annual meeting of the American
Sociological Association, San Francisco, California, September,
1982. Preparation of this paper was supported by National Insti-
tute of Child Health and Human Development Grant 1R01 HD 15730-01
to the University of Michigan, and by a grant from the Andrew W.
Mellon Foundation. We are extremely grateful to Carol Crawford
for her outstanding manuscript wordprocessing, Albert F. Anderson
for his superior computer programming and dedication, Joan Kahn
and Lisa Neidert for their excellent research assistance, and our
colleagues on the Michigan Comparative Fertility Project, Albert
I. Hermalin and George Wong, whose comments greatly improved the
quality of this paper.

Please do not quote or reproduce this material without the per-
mission of the authors.

(Draft. Comments invited.)

"PERMISSION TO REPRODUCE THIS
MATERIAL HAS BEEN GRANTED BY

W. M. Mason

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)."

I. INTRODUCTION

The real problems of contextual analysis concern the conceptualization of contextual effects, the kinds of data with which to estimate them, and the selection and implementation of appropriate statistical techniques. This paper focuses on detection. Specifically, we explicate an approach to contextual analysis based on the estimation and interpretation of a covariance component model with hyperparameters. The idea of this type of hyperparameter model is not new; it can be traced back at least as far as the Lindley and Smith (1972) contribution. However, the restricted maximum likelihood/Bayes (REML/Bayes) estimation procedure we describe is new, and is based on Mason and Wong (forthcoming).

The organization of the remainder of this paper is as follows. Section II describes our approach in simple and relatively informal terms, and Section III touches on estimation and computation. Section IV illustrates the specification, estimation, and interpretation of a covariance component model with hyperparameters, using an example concerning systematic variability across countries in the determination of individual-level fertility within countries. Our purpose in developing this example is to make the approach accessible to others interested in contextual analysis, but not in the algebraic and statistical complexities that comprise formal presentations.

II. MULTILEVEL FRAMEWORK

We begin by supposing that the researcher has data at two levels of observation, the individual (micro) and some higher

level (macro). The example we use for illustrative purposes will pertain to individuals (the micro level) and countries (the macro level), but others are easily conceived. Our fundamental assumption is that the micro values of the response variable in some way depend on context, and that the effects of the micro determinants may vary systematically as a function of context.

For simplicity we shall assume in the formal part of our discussion that there is one micro response variable (Y), one micro regressor (X), and one macro regressor (G). The empirical example developed in subsequent sections is considerably more complex. We define a micro equation identically for each context:

$$(1) \quad Y_{ij} = \beta_{0j} + \beta_{1j}X_{1ij} + \epsilon_{ij}$$

where $j = 1, \dots, J$ subscripts contexts and $i = 1, \dots, n_j$ subscripts micro observations within contexts. The total N for this setup is thus $\sum_j n_j$. Contexts are free to have differing numbers of micro observations.

The initial assumptions we make concerning this micro specification are as follows:

- (i) The micro model meets the usual rank condition using the data supplied for each context, with $\underline{X}$ fixed.
- (ii) $\epsilon_j \sim N(0, \sigma_j^2 I_{n_j})$.
- (iii) The J samples are independent, so that $\text{cov}(\epsilon_{j'}, \epsilon_j) = 0$ for $j' \neq j$.

Equations (2) and (3) describe the macro part of the model:

$$(2) \quad \beta_{0j} = \eta_{00} + \eta_{01}G_{1j} + \alpha_{0j} \quad ,$$

$$(3) \quad \beta_{1j} = \eta_{10} + \eta_{11}G_{1j} + \alpha_{1j} \quad .$$

We assume:

(iv) Each macro equation meets the usual rank condition using the data supplied for each context.

(v) $\alpha_{0j} \sim N(0, \gamma_{00})$, $\alpha_{1j} \sim N(0, \gamma_{11})$ and that

$\text{cov}(\alpha_{0j}, \alpha_{1j}) = \gamma_{01} \neq 0$ (where γ_{00} and γ_{11} are the macro error variances and γ_{01} is the macro error covariance).

(vi) The micro error variances are independent of the macro error variances and covariances.

Equations (2) and (3) represent the effects of contextual characteristic G on the two parameters of the micro model.

How are we to think about the sources of error in equations (2) and (3)? Standard textbook treatments assume that contexts are randomly sampled from an infinite population, and that there is no particular interest in the realized values of the micro coefficients (if they are present). In contrast, it is not the contexts but the errors in the macro equations (2)-(3) of our model that are uninteresting. To put this another way, once the systematic component has been removed from

$\beta_0 = (\beta_{01}, \dots, \beta_{0J})'$ and $\beta_1 = (\beta_{1j}, \dots, \beta_{1J})'$, their variability is strictly random. Alternatively, since the assumption of an infinite--or even very large--population of contexts from which to sample often seems unrealistic, we can appeal to the Bayesian perspective. From this standpoint, we can think of the α 's as exchangeable (Lindley and Smith, 1972; Barnett, 1973; Lindley and Novick, 1981): On the basis of our prior knowledge of contexts, we would not care if the α_{0j} 's and α_{1j} 's were permuted within equations (2) and (3), respectively. The exchangeability assumption does not require that we think of an infinite population of contexts.

We obtain a single-equation expression for the multilevel model by substituting equations (2) and (3) into equation (1):

$$(4) \quad y_{ij} = \eta_{00} + \eta_{01}G_{1j} + \eta_{10}X_{1ij} + \eta_{11}X_{1ij}G_{1j} \\ + [\alpha_{0j} + X_{1ij}\alpha_{1j} + \epsilon_{ij}]$$

The brackets in equation (4) enclose error terms which complicate the expression considerably, as they do also its estimation. This model is a mixed model, in statistical parlance, because it contains fixed coefficients (the η 's) and random coefficients (the α 's). Since the random coefficients are allowed to covary (over countries, in our illustrative example), this kind of model can also be called a covariance component model.

The model shown in equation (4) is quite general, in the sense that other models of potential interest can be derived from

it. The model becomes a familiar fixed effects model by setting α_{0j} and α_{1j} equal to zero:

$$(4a) \quad Y_{ij} = \eta_{00} + \eta_{01}G_{j1} + \eta_{10}X_{ij1} + \eta_{11}X_{ij1}G_{j1} + \epsilon_{ij} .$$

Equation (4a) assumes that coefficient variability across contexts is perfectly determined by contextual characteristic G , whereas equation (4) assumes the presence of error.

A random effects model is obtained by setting to zero all of the coefficients of X and G ($\eta_{01} = \eta_{10} = \eta_{11} = \alpha_{1j} = 0$):

$$(4b) \quad Y_{ij} = \eta_{00} + \alpha_{0j} + \epsilon_{ij} ,$$

This model includes no capacity for explaining variability in Y_{ij} , at either the micro or macro level, but it does include two sources of random variability in Y_{ij} . In contrast, the whole point of equation (4) is to employ both micro and macro variables to account for variability in Y_{ij} .

Our model is closely related to the random coefficient regression (RCR) model, as is seen by setting to zero the coefficients of G ($\eta_{01} = \eta_{11} = 0$). In that case, equation (4) simplifies to:

$$(4c) \quad Y_{ij} = \eta_{00} + \eta_{10}X_{ij1} + [\alpha_{0j} + X_{ij1}\alpha_{1j} + \epsilon_{ij}] ,$$

which retains the multiple sources of error, but loses the representation of a macro effect. This formulation assumes that contextual variability is uninteresting.

Finally, although equation (4) specifies cross-level interactions, we might wish to estimate a model which is additive with respect to context. This can be done in two ways. One is to assume that $\eta_{11} = 0$ and $\alpha_{1j} = 0$, for all j . Doing so gives:

$$(4d) \quad Y_{ij} = \eta_{00} + \eta_{01}G_{1j} + \eta_{10}X_{1ij} + [\alpha_{0j} + \epsilon_{ij}] .$$

This is an additive variance component model. Its use requires acceptance of the assumption that η_{10} , the effect of X , really is constant across contexts. The second alternative is to assume only that $\eta_{11} = 0$, which gives:

$$(4e) \quad Y_{ij} = \eta_{00} + \eta_{01}G_{1j} + \eta_{10}X_{1ij} + [\alpha_{0j} + X_{1ij}\alpha_{1j} + \epsilon_{ij}] .$$

This is an additive specification that allows for variability in the effect of X , but specifies that this variability is unsystematic--sweeping it into an error component. This is attractive when the variability of β_1 is not of interest to the investigator.

III. ESTIMATION

Estimation of mixed models can be complex, particularly when there is imbalance (the n_j differ). In the context of models for time-series of cross-sections, Mundlak (1978) derives an estimated generalized least squares (EGLS), or two-stage Aitken, estimator. Mason (1980) discusses the use of two-stage Aitken estimation procedures in the analysis of multilevel models. Here we describe an approach with more desirable

statistical properties--the REML/Bayes estimation procedure employed by Mason and Wong (forthcoming).

Parameter estimates using the REML/Bayes procedure are obtained as follows. Maximum likelihood estimates (MLEs) of the variance and covariance components are first obtained. These are substituted into a generalized least squares formula for the linear parameters (the η 's). Insofar as the variance and covariance components and the η 's are concerned, this procedure differs from EGLS in the way the variance and covariance components are estimated. The Mason-Wong procedure uses ML, whereas EGLS is based on nothing more than a consistent estimator for these parameters. The Bayes portion of the estimation procedure pertains to the GLS formula for the linear parameters, and is due to Lindley and Smith (1972) and Smith (1973), who used the Bayesian framework.

The computations for the REML/Bayes procedure are iterative, based on the EM algorithm (Dempster, Laird and Rubin, 1977). The EM algorithm uses the distinction between complete and incomplete data in a very general sense. In our application the incomplete data are the coefficients of the micro and macro variables and any cross-level interactions. The first cycle of the algorithm obtains starting values for the incomplete data and then computes the maximum likelihood estimates of targeted parameters as if the incomplete data were complete. In each succeeding cycle, the incomplete data are updated and new MLEs obtained. The cycles continue until the parameters to be

estimated by ML, and the incomplete data, converge to the desired tolerance.

In our application, the incomplete data are the linear parameters, and the parameters to be estimated by ML are $\phi = \{\sigma_1^2, \dots, \sigma_J^2; \Gamma\}$, where the elements of the matrix Γ are the variance and covariance components (the γ 's). The iterations proceed as follows. To obtain start values:

- (i) estimate the micro model, Eq. (1), for each context, using OLS;
- (ii) estimate the macro model, Eq.'s (2) and (3), using OLS.

Steps (i) and (ii) supply initial estimates of ϕ . Thereafter, the EM algorithm iterates over two main steps until convergence occurs:

- (iii) estimate the single-expression multilevel model, Eq. (4), based on current estimates of ϕ ;
- (iv) update ϕ and return to (iii).

A rudimentary computer program has been written to carry out these procedures. Our goal is a large, integrated package enabling researchers to estimate a wide variety of multilevel models.

Note that Mundlak's (1978) EGLS estimator amounts to the first iteration of the EM algorithm for our model, and recall that OLS estimates of equations (1)-(3) comprise the start values, along with the macro variance-covariance matrix computed

from the residuals of Eq.'s (2)-(3). Thus, we obtain OLS and EGLS estimates of the parameters of the multilevel model as by-products of the REML/Bayes estimation procedure, a feature of the Mason-Wong (forthcoming) approach to be exploited in Section IV of this paper.

IV. ILLUSTRATIVE EXAMPLE

We illustrate the multilevel model and REML/Bayes estimation procedure with a multilevel analysis of the determinants of children ever born in 15 less developed countries. The analysis was carried out under the aegis of the Michigan Comparative Fertility Project, which involves the collaborative efforts of Albert I. Hermalin and George Wong as well as the two authors of the present paper. Indeed, much of our methodological work on contextual analysis stems from the substantive needs of the Michigan Comparative Fertility Project.

A. Data and Variables

The micro data for the empirical illustration reported here are from the World Fertility Survey (WFS).¹ The WFS program sponsored sample surveys of standardized format in a large number of different countries, carried out within the span of a decade. Here, we use data for a single birth cohort of married respondents (age 40-44 in 1974) from 15 developing countries: Colombia (n = 309), Costa Rica (n = 341), Fiji

¹For an early description of the WFS see International Statistical Institute (1975). For recent results and perspectives, the Proceedings of the WFS Conference are extremely helpful (International Statistical Institute, 1981).

(n = 555), Guyana (n = 337), Indonesia (n = 838), Jamaica (n = 225), Jordan (n = 352), Kenya (n = 423), Korea (n = 711), Lesotho (n = 268), Malaysia (n = 749), Panama (n = 321), Peru (n = 605), Sri Lanka (n = 776), and Thailand (n = 482).² The micro variables are: children ever born (CEB), a count of the respondent's live births; wife's education (WED), coded as completed years of formal schooling; and wife's childhood residence (URBC), coded 1 if urban and 0 if rural.

The macro level in our illustration refers to countries. We include two variables which help define the context within which individual childbearing takes place: socioeconomic development; and family planning program effort. Gross national product per capita in 1965 (GNP65), expressed in 1970 U.S. dollars, measures development (World Bank, 1980). The family planning program effort measure is that used by Mauldin et al. (1978), and indexes on a 30 point scale a number of circa 1972 program characteristics. We defend the time referents of the macro variables vis-à-vis the micro variables elsewhere (Entwistle et al., 1982). Note that the different sources for the micro and the macro data help to minimize the possibility of correlated measurement error across levels of analysis.

²These were the countries available to us at the beginning. Since then, the list has expanded, and future work will be based on more countries. In the results to follow, individual countries will not be identified by name, in order to satisfy certain conditions associated with release of particular WFS micro data files to the Michigan Comparative Fertility Project.

B. Specification

The micro specification used in the example derives from a large structural equations model of fertility, contraceptive use patterns, and child mortality (Entwisle et al., 1982). Despite its complexity the structural equations model implies a simple reduced-form expression for CEB, in which CEB is a function of URBC, WED, WED², and URBC*WED. In contrast to the more typical, theoretically agnostic models of CEB, our micro model contains only terms in childhood residence and wife's education instead of a kitchen sink full of socioeconomic variables. These and only these are the exogenous variables in the Entwisle et al. (1982) micro model. Hence, our micro model is a complete specification, despite its brevity.³

Since there are only 15 countries across which to model coefficient variability, and there are four micro coefficients per country to be modelled, our empirical work omits URBC*WED from the micro specification in an attempt to stabilize parameter estimates. Thus, the model we estimate is

$$(6) \quad CEB_{ij} = \beta_{0j} + \beta_{1j}WED_{ij} + \beta_{2j}URBC_{ij} + \beta_{3j}WED_{ij}^2 + \epsilon_{ij},$$

where assumptions (i)-(iii) apply.

The macro model treats every parameter in the micro model as depending on the same two macro variables, GNP65 and FPE:

³This is subject to the qualification that the URBC*WED term has been omitted in the empirical example below.

$$\begin{aligned}
 \beta_{0j} &= \eta_{00} + \eta_{01}\text{GNP65}_j + \eta_{02}\text{FPE}_j + \eta_{03}\text{GNP65}_j * \text{FPE}_j + \alpha_{0j} \\
 (7) \quad \beta_{1j} &= \eta_{10} + \eta_{11}\text{GNP65}_j + \eta_{12}\text{FPE}_j + \eta_{13}\text{GNP65}_j * \text{FPE}_j + \alpha_{1j} \\
 \beta_{2j} &= \eta_{20} + \eta_{21}\text{GNP65}_j + \eta_{22}\text{FPE}_j + \eta_{23}\text{GNP65}_j * \text{FPE}_j + \alpha_{2j} \\
 \beta_{3j} &= \eta_{30} + \eta_{31}\text{GNP65}_j + \eta_{32}\text{FPE}_j + \eta_{33}\text{GNP65}_j * \text{FPE}_j + \alpha_{3j}
 \end{aligned}$$

Assumptions (iv)-(vi), appropriately generalized, apply to the equations of (6)-(7), or equivalently, to the multi-level specification corresponding to equation (4), namely

$$\begin{aligned}
 \text{CEB}_{ij} &= \eta_{00} + \eta_{01}\text{GNP65}_j + \eta_{02}\text{FPE}_j + \eta_{03}\text{GNP}_j * \text{FPE}_j \\
 &+ (\eta_{10} + \eta_{11}\text{GNP65}_j + \eta_{12}\text{FPE}_j + \eta_{13}\text{GNP}_j * \text{FPE}_j) * \text{WED}_{ij} \\
 (8) \quad &+ (\eta_{20} + \eta_{21}\text{GNP65}_j + \eta_{22}\text{FPE}_j + \eta_{23}\text{GNP}_j * \text{FPE}_j) * \text{URBC}_{ij} \\
 &+ (\eta_{30} + \eta_{31}\text{GNP65}_j + \eta_{32}\text{FPE}_j + \eta_{33}\text{GNP}_j * \text{FPE}_j) * \text{WED}_{ij}^2 \\
 &+ (\alpha_{0j} + \alpha_{1j}\text{WED}_{ij} + \alpha_{2j}\text{URBC}_{ij} + \alpha_{3j}\text{WED}_{ij}^2 + \epsilon_j) .
 \end{aligned}$$

Our illustrative analysis is based on a specification slightly simpler than equation (8). Preliminary results indicated that a number of terms could be omitted. In particular, it appears that a model in which the effect of education (described by linear and quadratic terms) has

unstructured variability across countries, but the effect of childhood residence varies systematically across countries, should fit the data about as well as the full model.* We nevertheless retain some terms to describe systematic variability in the education effects, despite their low descriptive levels of significance, for several reasons: 1) to see if the data are consistent with strong anticipations about the education effect; 2) to present the results for residence and education in a comparable manner; 3) to illustrate the complexity engendered by even a slight deviation from linearity in the specification of the micro model.

The multilevel model that we actually estimate in our empirical example is:

$$\begin{aligned}
 \text{CEB}_{ij} = & \eta_{00} + \eta_{01}\text{GNP65}_j + \eta_{02}\text{FPE}_j + \eta_{03}\text{GNP}_j*\text{FPE}_j \\
 & + (\eta_{10} + \eta_{11}\text{GNP65}_j + \eta_{12}\text{FPE}_j) * \text{WED}_{ij} \\
 (9) \quad & + (\eta_{20} + \eta_{21}\text{GNP65}_j + \eta_{22}\text{FPE}_j + \eta_{23}\text{GNP}_j*\text{FPE}_j) * \text{URBC}_{ij} \\
 & + (\eta_{30} + \eta_{31}\text{GNP65}_j + \eta_{32}\text{FPE}_j) * \text{WED}_{ij}^2 \\
 & + (\alpha_{0j} + \alpha_{1j}\text{WED}_{ij} + \alpha_{2j}\text{URBC}_{ij} + \alpha_{3j}\text{WED}_{ij}^2 + \epsilon_j) .
 \end{aligned}$$

*All tests of significance are based on ratios of coefficients to standard errors, or on the Wald statistic more generally.

It omits the $GNP * FPE * WED$ and $GNP * FPE * WED^2$ interactions of equation (8).

The multi-level model represented by equation (9) is inherently interactive, in the sense that the effect of each explanatory variable is inseparable from the effects of other explanatory variables. To present hypotheses about each of the partial derivatives, and to justify them convincingly, is beyond the scope of the present paper. We thus focus on a subset, in particular, the effects of education and residence.³ We give an intuitive grasp of these hypotheses via the following abstraction. Suppose the micro model contains a single explanatory variable, socioeconomic status (S), and that its effect is linear. We hypothesize that the S-effect on CEB switches from positive to negative as development proceeds, and that the S-effect is dampened as family planning programs become more effective and widespread. Because the hypothesized impact of family planning programs on the S-effect is to dampen regardless of the direction of the S-effect, there is a need to allow for the possibility of interaction between development and family planning program effort in the determination of variability in the S-effect. From this general approach can be

³The reason for choosing this subset is that if our hypotheses about the education and residence effects are incorrect, there is little reason for continuing the line of inquiry established in Entwisle et al. (1982). On the other hand, hypotheses about the partials for GNP65 and FPE could well receive no support from the data at hand, since these partials reflect macro associations between GNP65, FPE, and mean CEB, and there is no reason to believe that these macro associations are close to those found in much larger samples of countries. Recall that our sample of 15 countries is an accidental one.

derived specific hypotheses about the effects of childhood residence and education. Since this paper emphasizes the statistical framework, we defer further discussion to the next subsection.

C. Substantive Results

Table 1 presents the REML/Bayes estimates of the coefficients, and the ratios of coefficients to standard errors (hereafter "t-ratios"), of the multilevel model represented in Eq. (9).^{*} Note that we are explicitly describing the data in accordance with hierarchy; the presence of higher-order terms requires the presence of lower-order relatives for meaningful interpretation. For instance, the only education coefficient significant at conventional levels is that for the quadratic "main effect" ($\hat{\eta}_{30} = -.0280$), but we also retain the linear "main effect" ($\hat{\eta}_{10} = .1167$). A similar point holds for the "main effects" of GNP65, FPE, and GNP65*FPE. Only one of these effects is as large in absolute value as its standard error. The "main effects" of these variables are retained, however, because their interactions with URBC have high levels of descriptive significance.

Table 1 here

The effects of childhood residence and education on children ever born can not be read off from Table 1, because

^{*}These are based on 80 iterations of the EM algorithm, and a convergence criterion of .001 proportionate change in the estimates of successive iterations.

these effects depend on the level of GNP65 and FPE. For this reason, Table 2 presents the urban-rural contrast conditional on specified levels of GNP65 and FPE. The levels chosen are the minimum, median, and maximum for the sample of 15 countries. Tables 3 and 4 present the effect of education, conditional on macro characteristics, in a variety of ways.

Table 2 here

Table 2 displays substantial variability in the urban-rural contrast for CEB. In settings where there is no official family planning program, and where GNP per capita is the lowest in our data, urban women have an implied average of 1.2 children more than rural women. At the highest level of GNP per capita, urban women have an implied estimated average of 2.5 children less than rural women. This shift is consistent with our hypothesis that the linear effect of a socioeconomic variable would travel from positive to negative as socioeconomic development proceeds. It also indicates a sizeable impact for GNP--a shift from GNP per capita of \$65U.S. to \$593U.S. implies a change in the urban-rural contrast of about 3.7 children, in the absence of an official family planning program. A similar but weaker trend in the urban-rural contrast across levels of GNP holds for FPE evaluated at its median score (i.e., FPE = 16). At the highest FPE score, the trend disappears and even reverses slightly, with the urban-rural contrast negative at all levels of GNP65.

Turning next to the dependence of the URBC effect on FPE, we see that at the highest level of GNP per capita, the negative urban-rural contrast moves toward zero as family planning program strength increases. This is as hypothesized. On the other hand, for GNP65 at the median, the contrast increases slightly as FPE increases. Finally, at the lowest level of GNP per capita, the urban-rural contrast not only decreases with increasing FPE, it becomes negative at $FPE = 24$.

In all, the pattern of variability in the effect of childhood residence is about as expected.

We next consider the effect of education on children ever born. Because of the nonlinearity, it is difficult to define a simple summary of this effect. Table 3 displays the education effect conditional on the specific levels of GNP65 and FPE used earlier, and includes three pieces of information for each pair of GNP65 and FPE values: the conditional partial derivative with respect to education, the conditional inflection point (in years of schooling), and a summary measure of the derivative evaluated conditionally on GNP65 and FPE. Table 4 displays the education derivative conditional on GNP65, FPE, and three levels of education (0, 4, and 10 years). Table 4, which is derived from Table 3, is included only to help clarify the conditional derivatives in Table 3, and will not be discussed further.

Tables 3 and 4 here

Table 3 shows that only at the lowest level of GNP65 are the inflection points within the defined range of education. At

higher levels of GNP65, the inflection points are either negative or positive and large. Neither alternative is possible--there can not be negative years of schooling, and the top code for education is far less than 30.36 (the smallest positive inflection point). This indicates that the education effect is essentially linear at median and larger values of GNP65, and that the effect of education actually increases at the lower end of the education scale in contexts where GNP per capita is very low. This pattern is consistent with theoretical anticipations. In particular, Entwisle et al. (1982) argue that higher status families should be more likely than lower status families to innovate in fertility control (and thus have lower fertility), in contexts these authors label "traditional." Thus, the concave-downward effect of education for low GNP per capita is plausible and expected.

Looking next at the conditional first derivatives, it is clear that their "intercepts" become increasingly negative as GNP65 increases, and likewise for increases in FPE. The opposite trend holds for the "slope" of the conditional first derivatives: As GNP65 increases, the "slopes" become more positive, and similarly for increases in FPE. Although simpler than would have been the case were the $GNP65 \cdot FPE \cdot WED$ and $GNP65 \cdot FPE \cdot WED^2$ terms included in the model, the meaning of these patterns is still concealed. An additional summary measure has been included to help at this point.

For each derivative conditional on GNP65 and FPE, Table 3 reports the integral of the conditional derivative evaluated over

the range $[0, 10]$.⁷ These summaries can be thought of as areas describing the overall departure of the derivative from zero, or as conditional partial derivatives of the education effect for which the education component of the derivative has been removed by standardization. Either way, the summaries indicate that: The higher the GNP per capita, the more sizeable the education effect on CEB; and the higher the FPE score, the smaller the effect of education on CEB. This is consistent with our initial discussion of the hypothesized socioeconomic effect, to the extent it can be given the absence of the three-way interaction terms. That is, we observe that the education effect becomes more pronounced as GNP increases, and is attenuated by increases in family planning program effort.

The results for the education effect are consistent with anticipations, but nonetheless suggestive at best, since most of the coefficients involved have low levels of descriptive significance. Future work, based on a larger sample of countries, should be illuminating.

D. Comparisons of OLS, EGLS, and REML/Bayes

The EM algorithm is known to converge slowly. For our illustrative example, it took 80 iterations to obtain convergence using a fairly generous criterion of stability. In large problems this could be expensive, even though the iterations are

⁷Note that for GNP65 = \$61 the summary is reported as a sum of two numbers. For this level of GNP, there is a positive area and a negative area, which we have distinguished. Even if we take the absolute values of these components, their sum is the smallest in each column, and within the GNP65 = \$61 row, the sums of the absolute values decrease as FPE increases.

based on sufficient statistics after the initial pass of the data. Researchers might well ask whether preliminary model selection is worthwhile using OLS or EGLS, reserving the REML/Bayes calculations until most needed. The intermediate results provided by our computer program suggest that reliance on OLS for exploratory purposes could be highly misleading, and that even EGLS could prove misleading.

For comparative purposes, Table 5 pulls together the OLS, EGLS and REML/Bayes coefficients for equation (9), and also presents the t-ratios for the EGLS and REML/Bayes coefficients.^{*} Clearly the REML/Bayes and EGLS coefficients are closer to each other than either are to the OLS coefficients. In particular, the OLS coefficient for the "main effect" of GNP65 is much the largest of the three, and the OLS micro "main effects" as well as the cross-level coefficients are generally the smallest. Moreover, the REML/Bayes and EGLS coefficients always have the same sign, but the OLS signs occasionally differ from the REML/Bayes or EGLS signs. The OLS results clearly give a misleading assessment in this example.

Table 5 here

On the other hand, the coefficient estimates, taken alone, suggest that the EGLS coefficients may be good enough for

^{*}The OLS standard errors are meaningless numbers if the analyst believes that the micro parameters are not completely determined by the macro variables, as is the case here. Moreover, these particular coefficients are not even the correct fixed-effect estimates. For these reasons we omit OLS "t-ratios."

exploratory purposes. The t-ratios, however, do not bear this out. As can be seen in Table 5, analysts of the EGLS and REML/Bayes estimates would reach different conclusions about the statistical significance of the explanatory variables. The REML/Bayes results allow the inference that all of the terms involving childhood residence have high significance levels; the EGLS estimates do not. In general, reliance on EGLS for these data and variables would lead to a specification which differed but little from a random coefficient regression specification, whereas the REML/Bayes estimates clearly indicate systematic contextual variability in the effect of childhood residence, at a minimum.

To see why convergence required so many iterations in this example, we must look at the micro error variance estimates (the $\hat{\sigma}_j^2$) and the estimates of the variance-covariance components ($\hat{\Gamma}$), which are presented in Tables 6 and 7. The OLS, EGLS, and REML estimates of the micro error variances shown in Table 6 are highly similar to one another. The covariance components must therefore drive the iterations. As shown in Table 7, OLS and EGLS estimates of Γ are fairly similar (a major exception being γ_{02}). There is, however, a big difference between the REML and other estimates of γ_{02} , γ_{22} , and to a lesser extent γ_{12} . Evidently, changes in the estimates of the covariance components can have a major impact on the computed standard errors of the coefficients.

Tables 6 and 7 here

A final comparison of OLS, EGLS and REML/Bayes concerns within-country micro regressions. The OLS results, obtained separately for each country, can look quite different from the REML/Bayes (posterior) results, which borrow information from other countries in the estimation of the micro relationship for any one country. Table 8 presents the micro coefficients for four countries in which there are marked differences between the OLS and REML/Bayes estimates.

Table 8 here

As Table 8 shows, in each of the four countries, the REML/Bayes estimate of the urban-rural contrast in childhood residence (β_{2j}) is strikingly different from the OLS estimate. There are substantial differences, as well, in the education effect. In particular, in country A, OLS describes a concave-downward relationship (inflection point at 2.1 years of schooling), whereas REML/Bayes describes monotonicity (inflection point at $(-.3 \text{ years})$). In country B, OLS describes an implausible, increasingly positive education effect (inflection point at $-15.6 \text{ years of schooling}$), whereas the REML/Bayes pattern is a well-behaved nonmonotonic one. In country C, OLS traces a concave-downward education effect (inflection point at $.3 \text{ years of schooling}$), whereas REML/Bayes traces a monotonically decreasing effect. Finally, in country D, OLS describes a U-shaped education effect (inflection point is a minimum, at $3.9 \text{ years of schooling}$), whereas REML/Bayes describes a more

plausible concave-downward pattern (inflection point is a maximum, at 2.3 years of schooling). Thus, above and beyond the existence of differences between the OLS and REML/Bayes estimates, the OLS-estimated effects are implausible in two of these countries, whereas the REML/Bayes estimated effects are plausible in all four. This is striking, and will be the subject of further investigation.

V. CONCLUSION

Multilevel analysis using the covariance component framework is an advance in contextual analysis. With care and rigor it is possible to move beyond the long standing debate concerning the possibility and meaning of contextual analysis. Whether contextual effects exist for a given substantive problem is one matter; their estimation is another. Researchers may be hard pressed to find such effects unless they use efficient and appropriate estimation techniques in conjunction with appropriate theoretical reasoning. Our illustration embodies these points: Reliance on OLS would have yielded dust; EGLS would have done a little better, but not much; REML/Bayes extracted results which otherwise would have been invisible.

REFERENCES

- Barnett, V.
1973 Comparative Statistical Inference. New York: Wiley.
- Dempster, A. P., N. M. Laird, and D. B. Rubin
1977 "Maximum likelihood from incomplete data via the EM algorithm" (with discussion). Journal of the Royal Statistical Society, Series A 140:1-38.
- Entwisle, B., Hermalin, A. I., and W. M. Mason
1982 Socioeconomic Determinants of Fertility Behavior in Developing Countries: Theory and Initial Results. Report to National Academy of Sciences.
- International Statistical Institute
1975 The World Fertility Survey: The First Three Years, January 1972-January 1975. The Hague-Voorburg: Netherlands.
- 1981 World Fertility Survey Conference 1980: Record of Proceedings (3 volumes). Voorburg: Netherlands.
- Lindley, D. V. and A. F. M. Smith
1972 "Bayes estimates for the linear model" (with discussion). Journal of the Royal Statistical Society, Series B 34:1-41.
- Lindley, D. V. and M. R. Novick
1981 "The role of exchangeability in inference." The Annals of Statistics 9:45-58.
- W. M. Mason
1980 Some Statistics and Methodology of Multi-Level Analysis. Report to World Fertility Survey.
- W. M. Mason and G. Wong
forth-coming "Specification and estimation of a two-level hierarchical Bayesian model with application to World Fertility Survey data." Ann Arbor, Michigan: The Population Studies Center of The University of Michigan.
- Mauldin, W. P., B. Berelson, and Z. Sykes
1978 "Conditions of fertility decline in developing countries." Studies in Family Planning 9:89-147.
- Mundlak, Y.
1978 "Models with variable coefficients: integration and extension." Pp. 483-509 in P. Mazodier (ed.), The Econometrics of Panel Data (Number 30-31, April-September, of Annals de l'INSEE). Paris: Institut National de la Statistique et des Etudes Economiques.

Smith, A. F. M.

1973 "A general Bayesian linear model." Journal of the
Royal Statistical Society, Series B 35:67-75.

World Bank

1980 World Development Report. New York: Oxford University
Press.

Table 1. Children Ever Born Regressed on Micro and Macro Variables, and Their Interactions, Using a Mixed Model REML/Bayes Estimation Procedure

Variable	Coefficient Symbol	Coefficient Estimate	Ratio of Coefficient to Standard Error (t-ratio)
Intercept	η_{00}	6.9546	7.30
GNP65	η_{01}	.0043	1.19
FPE	η_{02}	-.0533	-.88
GNP*FPE	η_{03}	-.000007	-.04
WED	η_{10}	.1167	.74
WED*GNP	η_{11}	-.0007	-1.40
WED*FPE	η_{12}	-.0015	-.18
URBC	η_{20}	1.6652	2.49
URBC*GNP65	η_{21}	-.0070	-2.80
URBC*FPE	η_{22}	-.0946	-2.37
URBC*GNP65*FPE	η_{23}	.0003	2.14
WED ²	η_{30}	-.0280	-2.41
WED ² *GNP65	η_{31}	.00004	1.33
WED ² *FPE	η_{32}	.0004	.67

Source: Married women age 40-44 in 1974, in 15 World Fertility Survey countries. See text for details.

Table 2. The Childhood Residence Contrast Conditional on Gross National Product and Family Planning Program Effort

GNP Per Capita	<u>Family Planning Program Effort Score</u>		
	0	16	24
\$61	1.238	.017	-.593
\$284	-.323	-.473	-.548
\$593	-2.486	-1.153	-.487

Source: Table 1.

Note: Childhood residence is measured by URBC, for which urban residence = 1 and rural residence = 0.

Table 3. Partial Derivative of the Education Effect on Children Ever Born, and Other Descriptive Information, All Conditional on GNP65 and FPE

	Family Planning Program Effort Score		
	0	16	24
<u>GNP65 = \$61</u>			
Derivative	.074 - .051WED	.050 - .038WED	.038 - .032WED
Inflection point	1.45	1.32	1.19
Summary measure	.05-1.86 = -1.81	.03-1.43 = -1.40	.02-1.24 = -1.22
<u>GNP65 = \$284</u>			
Derivative	-.082 - .033WED	-.106 - .020WED	-.118 - .014WED
Inflection point	-2.48	-5.30	-8.43
Summary measure	-2.47	-2.06	-1.88
<u>GNP65 = \$593</u>			
Derivative	-.298 - .009WED	-.322 + .004WED	-.334 + .011WED
Inflection point	-33.11	80.50	30.36
Summary measure	-3.43	-3.02	-2.81

Source: Table 1.

Note: Where the "summary measure" is reported as a sum, the first component of the sum is the positive area; the second is the negative area. Thus, the total integral is evaluated from 0 years of schooling to the inflection point, and from the inflection point to 10 years of schooling. For additional details, see the discussion in the text.

Table 4. The Partial Derivative of the Education Effect on Children Ever Born, Conditional on GNP65, FPE and Education

	Family Planning Program Effort Score		
	0	16	24
<u>GNP65 = \$61</u>			
Years of school = 0	.074	.050	.038
Years of school = 4	-.130	-.102	-.090
Years of school = 10	-.436	-.330	-.282
<u>GNP65 = \$284</u>			
Years of school = 0	-.082	-.106	-.118
Years of school = 4	-.214	-.186	-.174
Years of school = 10	-.412	-.306	-.258
<u>GNP65 = \$593</u>			
Years of school = 0	-.298	-.322	-.334
Years of school = 4	-.334	-.306	-.290
Years of school = 10	-.388	-.282	-.224

Table 5. OLS, EGLS, and REML/Bayes Coefficients, with EGLS and REML/Bayes t-ratios, for the Children Ever Born Regression

Variable	Coefficient Symbol	Estimator		
		OLS	EGLS	REML/Bayes
Intercept	η_{00}	6.5236	6.8422 (7.47)	6.9546 (7.30)
GNP65	η_{01}	.0061	.0048 (1.37)	.0043 (1.19)
FPE	η_{02}	-.0145	-.0446 (-.76)	-.0533 (-.88)
GNP*FPE	η_{03}	-.00015	-.00004 (-.20)	-.000007 (-.04)
WED	η_{10}	.0309	.0891 (.53)	.1167 (.74)
WED*GNP	η_{11}	-.0005	-.00065 (-1.32)	-.0007 (-1.4)
WED*FPE	η_{12}	.0008	-.00046 (-.05)	-.0015 (-.18)
URBC	η_{20}	.4745	1.3706 (1.65)	1.6652 (2.49)
URBC*GNP65	η_{21}	-.0032	-.0063 (-2.03)	-.0070 (-2.8)
URBC*FPE	η_{22}	-.0313	-.0788 (-1.55)	-.0946 (-2.37)
URBC*GNP65*FPE	η_{23}	.00009	.0002 (1.18)	.0003 (2.14)
WED ²	η_{30}	-.0124	-.0230 (-1.58)	-.0280 (-2.41)
WED ² *GNP65	η_{31}	.00001	.000031 (.82)	.00004 (1.33)
WED ² *FPE	η_{32}	-.000075	.00021 (.29)	.0004 (.67)

Note: The OLS coefficients are not those of the fixed-effects model. See the text discussion of the EM algorithm for clarification.

Table 6. Alternative Estimates of the Micro Disturbance Variances (σ_j^2)

Country	Estimator		
	OLS	EGLS	REML/Bayes
1	14.01	14.02	14.09
2	13.22	13.24	13.23
3	9.54	9.54	9.55
4	9.00	9.02	9.07
5	7.70	7.69	7.68
6	10.72	10.72	10.72
7	9.07	9.07	9.07
8	10.93	10.90	10.89
9	8.73	8.70	8.69
10	9.02	9.02	9.03
11	10.20	10.19	10.21
12	11.06	11.05	11.05
13	8.03	8.03	8.03
14	3.53	3.52	3.52
15	9.52	9.50	9.50

Table 7. Alternative Estimates of Γ , the Elements of Which are the Macro Error Variances and Covariances

	α_0	α_1	α_2	α_3
<u>OLS</u>				
α_0	1.172			
α_1	-.177	.063		
α_2	.041	-.016	.269	
α_3	.008	-.004	-.001	.0004
<u>EGLS</u>				
α_0	1.269			
α_1	-.170	.057		
α_2	-.071	-.017	.228	
α_3	.008	-.003	-.0001	.0003
<u>REML</u>				
α_0	1.312			
α_1	-.161	.057		
α_2	-.125	-.024	.097	
α_3	.007	-.003	.003	.0002

Note: The elements of $\hat{\Gamma}$ are referred as $\hat{\gamma}$'s in the text.

Table 8. OLS, and REML/Bayes Estimates of the Micro Effects in Four Selected Countries

Variable	Coefficient Symbol	Estimator	
		OLS	REML/Bayes
<u>Country A</u>			
Intercept	β_0	7.64	7.83
WED	β_1	.22	-.01
URBC	β_2	-.16	-.73
WED ²	β_3	-.05	-.02
<u>Country B</u>			
Intercept	β_0	8.09	8.05
WED	β_1	.13	.28
URBC	β_2	-1.10	.07
WED ²	β_3	.00	-.04
<u>Country C</u>			
Intercept	β_0	6.95	6.90
WED	β_1	.02	.00
URBC	β_2	-1.20	-.68
WED ²	β_3	-.02	-.02
<u>Country D</u>			
Intercept	β_0	5.62	5.54
WED	β_1	-.05	.15
URBC	β_2	.12	1.41
WED ²	β_3	.01	-.03